Application of Pattern Recognition Techniques for MathE Questions Difficulty Level Definition

Beatriz Flamia Azevedo[1,2](✉), Roberto Molina de Souza[3], Maria F. Pacheco[1,2], Florbela P. Fernandes[1], and Ana I. Pereira[1,2]

[1] Research Centre in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, 5300-252 Bragança, Portugal
{beatrizflamia,pacheco,fflor,apereira}@ipb.pt

[2] Laboratório Associado para a Sustentabilidade e Tecnologia em Regiões de Montanha (SusTEC), Instituto Politécnico de Bragança, 5300-253 Bragança, Portugal

[3] Federal University of Technology - Paraná, 80230-901 Curitiba, Brazil
rmolinasouza@utfpr.edu.br

Abstract. Active learning is a modern educational strategy that involves students in the learning process through diverse interactive and participatory activities. The MathE platform is an international online platform created to support students and lecturers in the Mathematics teaching and learning process. This platform offers a tool to aid and engage students, ensuring new and creative ways to encourage them to improve their mathematical skills. The study proposed in this paper refers to a comprehensive investigation of the patterns that may exist within the set of questions available on the MathE platform. The objective is to investigate how to evaluate the student's opinions about the question's difficulty levels based on the variables extracted from student answers collected through surveys applied among the platform's users. Moreover, a comparative study between variables is performed using correlation and hypothesis tests. Furthermore, based on the results obtained for samples of different sizes, it was possible to define the most appropriate number of answers that should be considered to categorize the question's difficulty level. The results demonstrated that the variables extracted could be used to carry out the question level, and 30 answers are the most appropriate number of questions that must be used to categorize the question level.

Keywords: active learning · e-learning · higher education · Mathematics

This work has been supported by FCT Fundação para a Ciência e Tecnologia within the R&D Units Project Scope UIDB/05757/2020, UIDP/05757/2020, SusTEC LA/P/0007/2021 and Erasmus Plus KA2 within the project 2021-1-PT01-KA220-HED-000023288. Beatriz Flamia Azevedo is supported by FCT Grant Reference SFRH/BD/07427/2021.

A. I. Pereira et al. (Eds.): OL2A 2023, CCIS 1981, pp. 300–315, 2024.
https://doi.org/10.1007/978-3-031-53025-8_21

1 Introduction

We live in a time characterized by extraordinary and rapid changes. There is a continuous emergence and evolution of new knowledge, tools, and methods for practicing and communicating mathematics, even more regarding digital tools and artificial intelligence resources. The importance of understanding and utilizing mathematics in everyday life and the professional sphere has never been greater and will continue to grow. In this dynamic world, individuals with mathematical proficiency will have significantly enhanced opportunities and choices for shaping their futures. A solid foundation in mathematics opens doors to productive pathways, while a lack of mathematical competence limits those opportunities. All students should be given the opportunity and necessary support to learn meaningful mathematics with depth and comprehension [11].

Although mathematics is a cornerstone of many higher education courses and essential for understanding the world, comprehending its concepts, even basic ones, is seen by many as an insurmountable hurdle. While some people have a positive view of mathematics, considering it a fascinating and challenging discipline, most individuals report difficulty grasping its concepts, finding it abstract and uninteresting, and feeling intimidated and often incapable of understanding the calculations and formulas involved. These negative complaints may be linked to many individuals who have had negative experiences in mathematics education, such as a lack of adequate support or non-engaging and demotivating teaching approaches, contributing to a negative perception of the subject.

One way to change people's negative views about Mathematics is by offering an education based on active learning methods. Active learning is a student-centered pedagogical technique that promotes student learning more effectively than traditional or lecturer-centered approaches [6,8]. The fundamental role of active learning is to encourage students to participate in their education by analyzing, debating, researching, and producing, either in groups or individually. In this way, the student is no longer just a listener and becomes an active participant in the learning ecosystem [5].

The development of learning capabilities and the learning process is strongly dependent on the active involvement of students in their education [1]. Case study research shows that active teaching strategies increase lecture attendance, engagement, students' acquisition of expertise on the discipline and engagement improves students' capabilities [1,5,7,9]. Besides, students who are trained through active learning methodologies express a high level of satisfaction [1,5].

Thereby, the implementation of active learning methods involves challenges, and the incorporation of digital educational tools in classes is one way to support lecturers and students throughout the process. In this sense, the MathE platform (mathe.pixel-online.org) was created to provide not only in-class resources, but also an alternative way to teach and study Mathematics, alone, in groups, in or outside the classroom.

Since the platform was created, several works have been carried out to improve the platform, aiming at refining the platform system to offer a customized platform that meets the needs of its users [2,4]. In [4], it was identified that dividing the difficulty level of the questions into basic and advanced was not enough to categorize the questions according to the student's needs since the error rate in questions considered basic was usually very high. In addition, since many students had poor performance on the basic level questions, they rarely attempted to answer the advanced level ones, causing a certain demotivation in using the platform. Thus, the work [3] investigated the classification of questions into different difficulty levels using clustering algorithms through hierarchical and partitioning techniques. To categorize the questions through clustering, only the question's success rate was considered as an input variable for the clustering algorithms. However, since this rate is calculated using a binary output (0 - incorrect answer or 1 - correct answer), the success rate of the questions tends to converge to 0.5 over time, making it impractical to use it to define the question's level.

To define the questions' level, weighing the students' and lecturers' opinions about the question's difficulty is important. Besides, based on previous research, it is known that for some question marked as easy for a lecturer is considered difficult for the student, and vice-versa. To explore the student's opinions about the question's difficulty level, a more in-depth study is proposed in this work. Thereby, the occurrence of correct and incorrect answers for each question is evaluated to identify question patterns and classify them into different difficulty levels. First, a temporal study is proposed to analyze the number of correct and incorrect answers for each question over time. After that, a comparative study is conducted among the obtained variables, considering different answers quantities. These results will help define new variables to address the issues detected in previous works, and will also be crucial in identifying the most appropriate number of answers representing the student's perspective about the question classification.

This paper is organized as follows. After the introduction, Sect. 2 describes the main functionalities and resources available at the MathE platform. After that, Sect. 3 presents the methodologies proposed to extract the parameters and compare the data sample. The database utilized is defined in Sect. 3.3. Section 4 presents the results and the discussion of the work. Finally, Sect. 5 concludes the paper by establishing the future direction for this work.

2 MathE Platform

The MathE platform (mathe.pixel-online.org) is an online educational system designed to assist students who face challenges in learning college-level mathematics, as well as those seeking to enhance their understanding of a wide range of mathematical topics, all at their own pace. This platform provides free access to various resources, including videos, exercises, practice tests, and pedagogical materials, covering various areas of mathematics taught in higher education courses. Additionally, MathE maintains a presence on YouTube and social media platforms such as Facebook and Instagram.

Until May 2023, the platform has already attracted participation from 109 lecturers and 1435 students from 14 nationalities. Its current structure consists of three sections, as demonstrated at Fig. 1: **Student's Assessment**, which encompasses topic-specific multiple-choice questions categorized into two difficulty levels (basic and advanced) predetermined by lecturers associated with the platform; **MathE Library**, an extensive collection of valuable and diverse materials about the covered topics and subtopics, including videos, lessons, exercises, training tests, and other formats; and **Community of Practice**, a virtual space that facilitates interaction and collaboration among lecturers and students, allowing them to work together towards shared objectives and fostering a robust networked community.

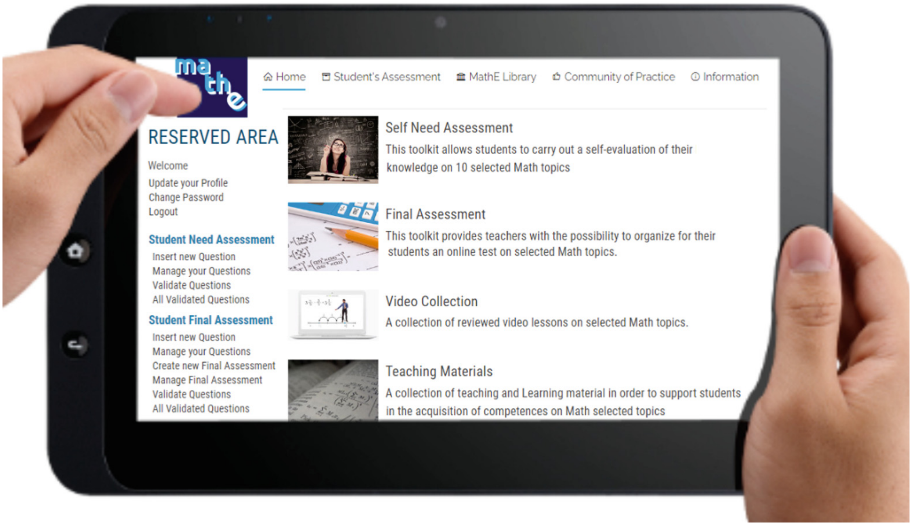

Fig. 1. MathE platform illustration.

MathE includes fifteen topics in Mathematics, among the ones that are in the classic core of graduate courses: Linear Algebra (5 subtopics: Matrices and Determinants, Eigenvalues and Eigenvectors, Linear Systems, Vector Spaces, and Linear Transformations), Fundamental Mathematics (2 subtopics: Elementary Geometry and Expressions and Equations), Graph Theory, Differentiation (including 3 subtopics: Derivatives, Partial Differentiation, Implicit Differentiation and Chain Rule), Integration (5 subtopics: Integration Techniques, Double Integration, Triple Integration, Definite Integrals, and Surface Integrals), Analytic Geometry, Complex Numbers, Differential Equations, Statistics, Real Functions of a Single Variable (2 subtopics: Limits and Continuity, and Domain, Image and Graphics), Probability, Optimization (2 subtopics: Linear Optimization and Nonlinear Optimization), Real Functions of Several Variables (2 subtopic: Limits and Continuity, Domain, Image and Graphics), and Numerical Methods. It is essential to mention that the platform's content is constantly updated, and other topics and subtopics may be created whenever necessary.

3 Methodology

As previously mentioned, the variable question success rate is not enough to categorize the level of the questions since it over time tends to be 0.50 for all questions. Then, this section describes the methods applied in this work to obtain other parameters through the information of the type of answer (correct or incorrect) provided by the students. Moreover, these parameters are compared considering different data samples.

3.1 Obtaining Parameters

An analysis of the distribution of correct and incorrect answers to questions over time is proposed to obtain new parameters from the answers collected. However, there is no information regarding the data and time when the questions were answered on the platform. On the other hand, obtaining the sequence of entries for the students' answers is possible, allowing for determining the order of the solutions obtained for each question. Considering this, Fig. 2 aims to illustrate the proposed concepts by analyzing three questions represented by colored asterisks: question 1 - green, question 2 - blue, and question 3 - red.

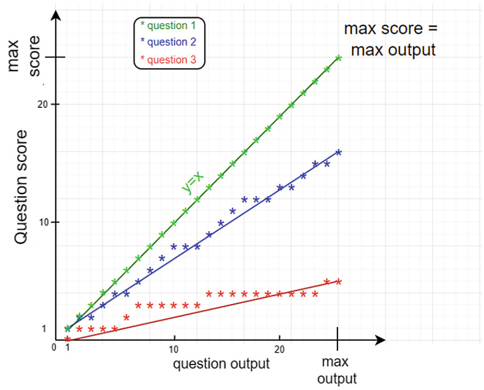

Fig. 2. Example of question answers over time (Color figure online).

Since correct answers are represented by the value 1 and incorrect answers by 0, 1 is added to the current question score for each correct answer. Similarly, for each incorrect answer, 0 is added. The higher the number of correct answers, the higher the score of the question, with the maximum score obtained if all answers to the question are correct, as illustrated by question 1 (green). Thereby, the x-axis represents the chronological order of the answers to the questions over time, where a coordinate closer to $x = 1$ indicates an earlier response recorded by the platform. Meanwhile, the y-axis represents the score obtained by the question, considering a certain number of outputs. This allows for an analysis of the answers over time. Consider the example of question 3 (red), where the answers vector is represented by $\mathbf{v3} = [0, 1, 0, 0, 0, 1, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0, 0, 0, 1, 0]$, which means that the first answer to question 3 was incorrect (0), the second answer was correct (1), the third, fourth, and fifth responses were incorrect, and so on.

With this data distribution for each question, it is possible to obtain the line that describes the points on the graph through linear interpolation [13], where $P(x)$ is the interpolating polynomial and m represents the angular coefficient of the line obtained by P, as shown in Eqs. 1 and 2, respectively. Here, y_1 and y_0 are the scores of the questions, and the number of questions' outputs is x_1 and x_0.

$$P(x) = y_0 + m(x - x_0) \tag{1}$$

$$m = \frac{y_1 - y_0}{x_1 - x_0} \tag{2}$$

When the maximum score is achieved (as in question 1, in the previous figure), we have $y = x$. In the other cases, we have variations in the slope of this line. Therefore, through the linear equations of each question, it is possible to evaluate the angular coefficient m of the line. Thus, questions with more incorrect answers will have smaller angular coefficients and are considered more difficult than those with larger coefficients.

Through this method, it is also possible to calculate the cumulative scores (cs) for each question so that questions with lower cumulative scores are more difficult than those with a higher cs value. In Fig. 2, question 1 has an angular coefficient of $m = 1$, and the highest possible cumulative score cs for the maximum number of outputs considered since all the answers were correct. This indicates that question 1 can be considered an easy-level question. On the other hand, question 2 has an intermediate value of the angular coefficient, suggesting it is an intermediate-level question. In turn, question 3 has the lowest angular coefficient classifying it as a difficult-level question. Considering this method, it is possible to obtain two new variables to analyze the question levels, the cumulative score cs, and the angular coefficient m of each question.

3.2 Sample Data Comparison Methods

In order to compare the variables extracted by the methodology proposed at Subsect. 3.1, two analyses are performed: the correlation analysis and the Hypothesis test. The correlation evaluates if the variables cs and m could replace the success rate variable without compromising the results. The Hypothesis test is used to evaluate the sample's most appropriate size, which is the number of answers, into the question categorization.

Correlation Analysis. To compare the correlation between the obtained parameters, the Pearson method is applied. The method aims to check if the variables cs and m when used instead of the success rate does not imply significant changes in the results for classifying the difficulty level of the questions.

Hypothesis Test. To evaluate the most suitable sample size of data to determine the difficulty level of the question, a hypothesis test is considered. In this case, it is intended to investigate whether a parameter collected using different sizes samples has equal means or not. Thereby, the t-Student test [10] is used to determine if there is a significant difference between the means μ_1 and μ_2 composed of 30 and 50 answers per question, respectively. So, the two hypotheses are considered:

$$\begin{aligned} H_0 &: \mu_1 = \mu_2 \\ H_1 &: \mu_1 \neq \mu_2 \end{aligned} \tag{3}$$

in which, H_0 assumes that there is no difference between μ_1 and μ_2, whereas H_1 affirms there is difference between μ_1 and μ_2.

Thereby, to perform the t-test a significance level of 0.05 is considered. So, if the p-value found is less than the chosen significance level ($p < 0.05$), the null hypothesis is rejected, suggesting a difference between the means of the samples compared. But, if the p-value is greater than or equal to the significance level, the null hypothesis is not rejected suggesting that there is not enough evidence to support a difference between the means of the samples compared [10]. It is essential to point out that before applying the t-Student test, it is necessary to verify the normality of the data distribution. For this, the Kolmogorov-Smirnov test [10] was used, also with a significance level of 0.05.

3.3 Data Base

Currently, the MathE platform has 1824 questions available, distributed between 15 topics and 24 subtopics, in which the subtopics Vector Space and Linear Transformation are the two most used subtopics of the platform. Both of them belong to the Linear System topic. Table 1 describes the data considered per subtopic. So, column *N. questions* shows the number of questions considered in this work; column *N. Answers* presents the total number of answers obtain, and the column *N. students* represents the number of different students that answered the questions since one question could be answered more than one time for the same student. However, for applying the methods used in this study, only the questions that received at least 50 responses were considered, which reduced that number of questions to 21 Linear Transformation questions and 24 questions at the Vector Space subtopic.

Table 1. Data base description.

Subtopic	*N. questions*	*N. Answers*	*N. students*
Linear Transformation	60	2067	31
Vector Space	61	2738	96

4 Results and Discussion

As previously mentioned, the information currently available on the MathE platform refers to the type of answer provided by the students (1 - correct or 0 - incorrect). Thus, this is an equiprobable sample space, with a success rate equal to 0.5, which makes this variable impracticable for classifying the questions in different difficulty levels. In this way, the need to extract other information from the available information was observed. Thus, with the methodology described in Sect. 3.1, it is possible to obtain the variables denominated by the cumulative score of the question (cs) and the angular coefficient (m). The variables obtained are analyzed through correlation and t-Student hypothesis methods. The results are presented below for the two most used subtopics of the platform: Linear Transformations (Sect. 4.1) and Vector Spaces (Sect. 4.2).

4.1 Linear Transformations Results

The Linear Transformations (LT) subtopic currently comprises 60 questions. However, only 21 questions have at least 50 answers collected since the MathE platform has been online. For this reason, these questions were selected to be analyzed since they have the minimum number of answers for applying the methods considered in this work.

Thereby, to provide a comparison between students' and lecturers' question level definition, the 21 questions of the Linear Transformation subtopic are classified into three groups by a lecturer, according to their knowledge. Each group means a difficult level, starting in the 1 - most basic level and 3 - most difficult level. After that to be better analyzed, three questions were randomly selected, one of each difficulty level provided by the lecturer. In this case, questions 287 is selected from level 1, questions 784 is selected from level 2, and questions 384 is selected from level 3. Figure 3 presents the profile of these questions, considering a maximum output equal to 50 (x-axis). And consequently, 50 is the maximum score to be achieved in case all answers provided are correct.

Based on the results presented at Fig. 3, the student's classification of the questions based on the question's success rate is equal to the classification provided by the lecturer for these three questions. But this is not always the case, sometimes while the lecturer indicates that the question 325 is basic, there were many incorrect answers, indicating that this question is not really basic from the student's point of view. However, it is important to highlight that they are some situations in which the classification of the student and lecturer are different. Such results demonstrate the reason for analyzing the two options: the

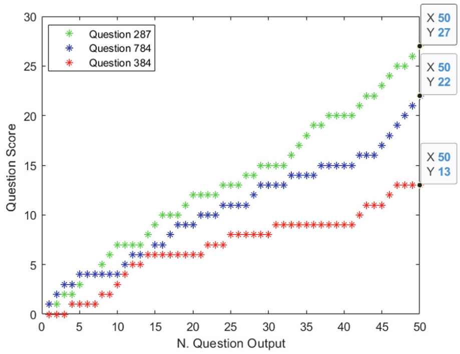

Fig. 3. Example of questions behavior 287, 784, and 384.

student's classification provided by the question success rate and the lecturer's classification based on their experience.

Considering the analysis proposed at Sect. 3.1, the linear interpolation was performed, and the parameter's cumulative score cs, angular coefficient m, and the success rate for each question were obtained. For this analysis, different quantities of answers were considered to analysis the behavior of consider different amount of data to extract the previously mentioned parameters. Thereby, Table 2 presents the data extracted for the three questions of Fig. 3. Considering different numbers of outputs (answers), it is 10, 20, 30, 40, and 50 for each question, it is noticed that the success rate, even in questions belonging to different difficulty level, tend to be close to 0.5 over time.

Table 2. Success rate for different numbers of outputs for the three LT questions.

Question ID	Level given by lecturer	Number of Output				
		N10	N20	N30	N40	N50
Question 287	1	0.90	0.60	0.50	0.50	0.54
Question 784	2	0.40	0.45	0.43	0.38	0.44
Question 384	3	0.30	0.30	0.26	0.23	0.26

On the other hand, Table 3 presents the values referring to the variable cumulative score, cs, also considering samples of different sizes. Besides the value, it is also presented the normalized value of the variable for a better understanding of the position of the question in the range of values between 0 and 1 and the definition of the most appropriate level of difficulty for the question.

Comparing Tables 2 and 3, there is a greater gap between the classes of the cs normalized values and success rate values. For the case of 50 answers, in Tables 2, we have a difference of 0.18 between the question of level 3 and level 2, and a difference of 0.1 between the question of level 2 and level 1; whereas, in Table 3,

Table 3. Cumulative score for different numbers of output for the three LT questions.

Question ID	Level given by lecturer	Number of Output				
		N10	N20	N30	N40	N50
Question 287	1	35 (0.63)	126 (0.69)	259 (0.63)	438 (0.63)	673 (0.68)
Question 784	2	33 (0.58)	105 (0.51)	217 (0.43)	359 (0.39)	539 (0.40)
Question 384	3	11 (0.04)	67 (0.18)	142 (0.08)	232 (0.00)	348 (0.00)

this difference is approximately 0.3 for the question of level 1 and 2, and 0.4 to questions of level 2 and 3.

Table 4 presents the three questions' angular coefficient m. As we can see, the m values are very similar to the success rate values, indicating that the variable cs is the most appropriate to replace the success rate variable at the questions level definition.

Table 4. Angular coefficient for different numbers of output for the three LT questions.

Question ID	Level given by lecturer	Number of Outputs				
		N10	N20	N30	N40	N50
Question 287	1	0.66	0.60	0.48	0.48	0.53
Question 784	2	0.33	0.42	0.41	0.36	0.42
Question 384	3	0.33	0.31	0.27	0.23	0.27

Considering that the MathE Platform is a world-class platform and through previous studies, a sample of 10 or 20 answers is too small to define the level of difficulty of the questions, that represent the opinion of student worldwide. So, a more deep analysis involving the sample of 30 and 50 answers is proposed. In this way, the Pearson correlation and the hypothesis test between the samples composed of 30 and 50 answers were performed.

The Pearson correlation was used to compare the correlation between the success rate, angular coefficient m, and cumulative score cs variables, for the 21 questions in the study. The result is shown in Table 5. As can be seen, there is a high correlation, over 0.8, between all variables independent of the number of outputs considered. This indicates that variables cs and m can substitute the success rate variable. However, as previously mentioned the m values are very similar to the success rate values. So, using the m variable, the tendency of the variable comes close to 0.5 is kept. Thus, the cs results are more expressive to the used in the question level definition.

Finally, the most appropriate size of the data was investigated. First, the questions were divided into 3 levels of equal sizes for each number of outputs

Table 5. Correlation between the variables of the LT.

N. Outputs	success rate & cs	cs & m	success rate & m
N30	0.93	0.89	0.99
N50	0.90	0.88	0.99

considered. In the case of the subtopic Linear Transformations, each level is composed of 7 questions, using the values of the variable cs. Thus, the first 7 questions with the lowest cs are at level 1, the following at level 2, and level 3 are the questions with the highest cs. This process was done for each sample independently. Since the lecturers classified the same questions, it is possible to compare the number of questions at the same level according to the lecturer's opinion with the classification provided by the students through the questions' answers. The results of this analysis are shown in Table 6. Each value in the table represents the number of questions at the intersection between one sample and another, 2 to 2 (Table 6a), and the similarity rate between the samples, in terms of question levels categorization (Table 6b). In other words, it means the number of questions that keeps at the same difficulty level in both numbers of output.

Table 6. Question levels comparison at LT subtopic.

(a) Number of questions at the intersection.

Sample	Lect.	N30	N50
Lect.	**21**	11	9
N30		**21**	17
N50			**21**

(b) Similarity rate between the levels.

Sample	Lect.	N30	N50
Lect.	**21**	0.52	0.43
N30		**21**	0.81
N50			**21**

As we can see, there are 17 questions that remain at the same levels, considering both a sample of 30 and 50 answers, which results in a similarity rate equal to 0.81. Regarding the classification of the lecturer and the students, we have 0.52 and 0.43 similarity rates between the classifications considering the samples composed of 30 and 50 answers. These results were already expected, given the knowledge of previous work that pointed to several divergences between the opinion of students and lecturers.

After checking whether the data follows a normal distribution, the hypothesis test was applied to compare the statistical means between the 30 and 50 samples. Through this, it is possible to define whether the observed sample differences are real or casual. It is important to clarify that the cs values were normalized to be compared by the hypothesis test approach. The results are shown in Table 7, in which the p-value model test equals 0.39. Considering a significance level equal to 0.05, the final decision considers not rejecting the null hypothesis (Sect. 3.2) since the p-value found are greater than 0.05. Therefore, there is no significant difference between the means of the two samples.

Table 7. t-Student test results for LT questions.

t	Degree of Freedom	Standard Deviation	p-value	Confidence Interval (95%)
−0.86	40	0.25	0.39	[−0.22, 0.09]

4.2 Vector Spaces Results

The Vector Spaces (VS) subtopic is composed of 61 questions, but only 24 had more than 50 answered computed, so only these were considered. The process applied to the Linear Transformations subtopic was replayed to the Vector Spaces subtopic. This way, a lecturer analyzed the 24 questions and categorized them into three groups. Figure 4 describes the profile of 3 questions randomly selected, 1 for each group, being the question 417 belongs to level 1; question 421 belongs to level 2; and question 453 belongs to level 3.

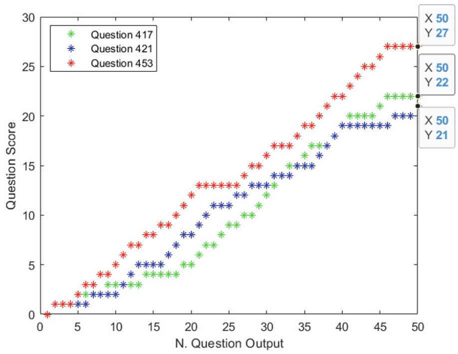

Fig. 4. Example of questions behavior 417, 421, and 453.

These questions were chosen to be better analyzed. So, Tables 8, 9 and 10 presents the value of the variable's success rate, the cumulative score cs, and the angular coefficient m, respectively, for the three questions.

Table 8. Success rate for different numbers of outputs for the three VS questions.

Question ID	Level given by lecturer	Number of Outputs				
		N10	N20	N30	N40	N50
Question 417	1	0.30	0.25	0.40	0.47	0.44
Question 421	2	0.40	0.45	0.40	0.37	0.36
Question 453	3	0.20	0.40	0.43	0.47	0.42

Considering the results of Table 8, the variation between the success rate of the three questions is quite small. In the case of 30 outputs, it is noteworthy that the three questions have a success rate of approximately 0.4. Similar results are obtained in Table 10.

Table 9. Cumulative score for different numbers of output for the three VS questions.

Question ID	Level given by lecturer	Number of Outputs				
		N10	N20	N30	N40	N50
Question 417	1	16 (0.28)	55 (0.27)	144 (0.34)	305 (0.43)	516 (0.46)
Question 421	2	21 (0.39)	82 (0.46)	180 (0.49)	315 (0.45)	483 (0.40)
Question 453	3	13 (0.21)	69 (0.37)	184 (0.49)	341 (0.52)	536 (0.50)

On the other hand, analyzing the Table 9 results, there is a more expressive difference between the values of the questions, even when we consider normalization.

Table 10. Angular coefficient for different numbers of output for the three VS questions.

Question ID	Level given by lecturer	Number of Outputs				
		N10	N20	N30	N40	N50
Question 417	1	0.33	0.26	0.41	0.49	0.44
Question 421	2	0.44	0.47	0.41	0.38	0.36
Question 453	3	0.22	0.42	0.45	0.49	0.42

Again, the correlation between the variables was evaluated. Thus, Table 11 presents the correlation between the three variables: success rate, cumulative score cs, and angular coefficient m, for the 24 Vector Spaces questions.

As occur in the Linear Transformations questions, both combinations of variables presented high correction, over 0.80, which indicates that the success rate variable could be replaced by the other two variables cs or m, without compromising the results. But, the success rate values are practically the same as the m values, so cs is the most appropriate variable for the question classification.

For the analysis of the most appropriate size of the data, it is, the number of questions to be used to define the cs values, the 24 Vector Spaces questions were categorized into 3 levels of equal size, in this case, 8 questions per level. Table 12 presents the results, considering the number of questions belonging to each sample's intersection, it is the same difficulty level, considering the lecturer classification and the different numbers of output (answers), 2 to 2 (Table 12a),

Table 11. Correlation between the variables of the VS.

N. Output	Success rate & cs	cs & m	Success rate & m
N30	0.84	0.81	0.99
N50	0.90	0.89	0.99

and the similarity rate between the questions difficult level (Table 12b). Again, a higher similarity between the sample of 30 outputs and the sample of 50 was found, equal to 0.83; while in the other cases, this similarity is not so higher, below equal to 0.46 for the comparison between the lecturer and 50 answers, and also between the sample of 30 and 50 answers.

Table 12. Question levels comparison at VS subtopic.

(a) Number of questions at the intersection.

Sample	Lect.	N30	N50
Lect.	**24**	11	11
N30		**24**	20
N50			**24**

(b) Similarity rate between the levels.

Sample	Lect.	N30	N50
Lect.	**24**	0.46	0.46
N30		**24**	0.83
N50			**24**

Finally, after the Kolmogorov-Smirnov test indicates that the data is normal, the t-Student test was performed considering the cs normalized variable for the samples composed of 30 and 50 answers. As presented in Table 13, the p-value found is 0.93. In this way, considering a significance level equal to 0.05, the final decision considers not rejecting the null hypothesis (Sect. 3.2) since the p-value found is greater than 0.05.

Table 13. t-Student test results for VS questions.

t	Degree of Freedom	Standard Deviation	p-value	Confidence Interval (95%)
−0.08	46	0.22	0.93	[−0.13, 0.12]

5 Conclusion

The MathE platform has been online since 2019. The necessity for some improvements on the platform has been identified in previous studies, such as question difficult level definition and the optimum way to reorganize the resources available on the platform [2–4]. The insights gained from prior research have guided the developers of the MathE platform in implementing intelligent features, leveraging optimization algorithms and machine learning techniques. These advancements will enable the platform to autonomously make personalized decisions tailored according to the needs of each user. This work investigated different variables to be used to define the question's difficulty level. Furthermore, the most appropriate number of answers to obtain the variable's information was also investigated.

Through previous works, it is known that the success rate variable is not enough to categorize the questions, taking into account that the information considered is a binary variable (1 - correct, and 0 - incorrect), the success rate of the answers to the questions is 0.5, and this is observed by computing responses over time. Thus, two new variables were extracted for each question: the angular coefficient m, and the cumulative score cs. Taking into account that the angular coefficient and the success rate have a high correlation and therefore present the same problem as the success rate variable, it is defined that the cumulative score variable is the most recommended to be used in the question classification task, and represent the students' opinions about the difficulty of each question. Regarding the number of samples to be used, although the sample 30 and 50 answers have no significance, it was established that 30 answers are the most appropriate value to obtain cs values and represent the students' opinions. Besides, from the literature, it is known that 30 samples are considered a minimal quantity to obtain trustful in the data analysis results [12].

From previous work, it is known that the opinion of students and lecturers, about the complexity of the questions is often divergent. While for a lecturer, a question may be very basic, for the student the same question may be considered very difficult, and vice-versa. Thus, it is important to give appropriate weight to the opinion of both students and lecturers, in order to obtain the most accurate classification. Thus, the focus of this work was the exploration of the best way to extract the students' opinions, which will be combined with the lecturer's opinion in future works. Thereby, the values of these opinion combinations will, in the future, be considered for the final classification of the questions.

References

1. Alhawiti, N.M.: The influence of active learning on the development of learner capabilities in the college of applied medical sciences: mixed-methods study. Adv. Med. Educ. Pract. **14**, 87–99 (2023)
2. Flamia Azevedo, B., Rocha, A.M.A.C., Fernandes, F.P., Pacheco, M.F., Pereira, A.I.: Evaluating student behaviour on the mathe platform - clustering algorithms approaches. In: Simos, D.E., Rasskazova, V.A., Archetti, F., Kotsireas, I.S., Pardalos, P.M. (eds.) Learning and Intelligent Optimization, LION 2022. LNS, vol. 13621, pp. 319–333. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-24866-5_24
3. Azevedo, B.F., Amoura, Y., Rocha, A.M.A.C., Fernandes, F.P., Pacheco, M.F., Pereira, A.I.: Analyzing the MathE platform through clustering algorithms. In: Gervasi, O., Murgante, B., Misra, S., Rocha, A.M.A.C., Garau, C. (eds.) Computational Science and Its Applications, ICCSA 2022 Workshops, ICCSA 2022. LNCS, vol. 13378, pp. 201–218. Springer, Cham (2022). https://doi.org/10.1007/978-3-031-10562-3_15

4. Azevedo, B.F., Pereira, A.I., Fernandes, F.P., Pacheco, M.F.: Mathematics learning and assessment using MathE platform: a case study. Educ. Inf. Technol. **27**, 1747–1769 (2021)
5. Deslauriers, L., McCarty, L.S., Miller, K., Callaghan, K., Kestin, G.: Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom. Proc. Nat. Acad. Sci. **116**(39), 19251–19257 (2019)
6. Dunkle, K.M., Yantz, J.L.: Intentional design and implementation of a "flipped" upper division geology course: improving student learning outcomes, persistence, and attitudes. J. Geosci. Educ. **69**(1), 55–70 (2021)
7. Gu, X., et al.: Active versus passive strategy in online creativity training: how to best promote creativity of students with different cognitive styles? Think. Skills Creat. **44**, 101021 (2022)
8. Indorf, J.L., et al.: Distinct factors predict use of active learning techniques by pre-tenure and tenured stem faculty. J. Geosci. Educ. **69**(4), 357–372 (2021)
9. Keiler, L.S.: Teachers' roles and identities in student-centered classrooms. Int. J. STEM Educ. **5**(1), 34 (2018)
10. Montgomery, D.C.: Design and Analysis of Experiments. Wiley, Hoboken (2017)
11. NCTM: Principles and Standards for School Mathematics, vol. 1. National Council of Teachers of Mathematics (NCTM) (2000)
12. Pagano, M., Gauvreau, K.: Principles of Biostatistics. Duxbury Press (2000)
13. Weisstein, E.W.: Linear interpolation. From MathWorld - A Wolfram Web Resource. http://mathworld.wolfram.com/LinearInterpolation.html